Support Recovery with Orthogonal Matching Pursuit in the Presence of Noise: A New Analysis

Jian Wang

Department of Electrical & Computer Engineering

Duke University

Durham, North Carolina, 27708, USA

E-mail: jian.wang@duke.edu

Abstract—**Support recovery of sparse signals from compressed linear measurements is a fundamental problem in compressed sensing (CS). In this paper, we study the orthogonal matching pursuit (OMP) algorithm for the recovery of support under noise. We consider two signal-to-noise ratio (SNR) settings: i) the SNR depends on the sparsity level K of input signals, and ii) the SNR is an absolute constant independent of K. For the first setting, we establish necessary and sufficient conditions for the exact support recovery with OMP, expressed as lower bounds on the SNR. Our results indicate that in order to ensure the exact support recovery of all K-sparse signals with the OMP algorithm, the SNR must at least scale linearly with the sparsity level K. In the second setting, since the necessary condition on the SNR is not fulfilled, the exact support recovery with OMP is impossible. However, our analysis shows that recovery with an arbitrarily small but constant fraction of errors is possible with the OMP algorithm. This result may be useful for some practical applications where obtaining some large fraction of support positions is adequate.**

Index Terms—**Compressed sensing (CS), orthogonal matching pursuit (OMP), restricted isometry property (RIP), signal-to-noise ratio (SNR), minimum-to-average ratio (MAR).**

I. INTRODUCTION

We consider the support recovery of a high-dimensional sparse signal from a small number of linear measurements. This is a fundamental problem in compressed sensing (CS) [1], [2] and has also received much attention in the fields of sparse approximation [3], signal denoising [4], and statistical model selection [5]. Let $\mathbf{x} \in \mathcal{R}^n$ be a K-sparse signal (i.e., $\|\mathbf{x}\|_0 \leq K \ll n$) and $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ ($m < n$) be the measurement matrix. The measurements are given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v} \tag{1}$$

where $\mathbf{v}$ is the corrupting noise. The goal of support recovery is to identify the support (i.e., the positions of nonzero elements) of the input signal $\mathbf{x}$ from the measurement vector $\mathbf{y}$. It is known that optimal support recovery requires an exhaustive search over all possible support sets of the sparse signal and hence is NP-hard [6]. For this reason, much attention has been drawn to computationally efficient approaches. In this paper we consider the orthogonal matching pursuit (OMP) algorithm for solving the support recovery problem. OMP is a canonical greedy algorithm for sparse approximation in signal processing [7], [8]. It is also known as greedy least square regression in statistics [5] and forward greedy selection in machine learning [9]. The principle of the OMP algorithm is

quite simple: it iteratively identifies the support of the sparse signal, by adding one index into the list at a time according to the maximum correlation between columns of measurement matrix and the current residual. There are several popular stopping rules for the OMP algorithm that can be implemented at minimal cost [10]:

i) Stop after a fixed number of iterations: $k = K$.
ii) Stop when the energy in the residual is small: $\|\mathbf{r}^k\|_2 \leq \epsilon$.
iii) Stop when no column in the measurement matrix is strongly correlated with the residual: $\|\mathbf{\Phi}'\mathbf{r}^{k-1}\|_\infty \leq \epsilon$.

See Table I for the mathematical description of a version of OMP. Both in theory and in practice, the OMP algorithm has demonstrated competitive performance [11].

Over the years, many efforts have been made to analyze the performance of OMP in sparse support recovery. In one line of work, probabilistic analyses have been proposed. Tropp and Gilbert showed that when the measurement matrix $\mathbf{\Phi}$ is generated iid at random, OMP can ensure the accurate recovery of every fixed K-sparse signal from noise-free measurements with overwhelming probability with [11]

$$m \geq cK \log n \tag{2}$$

for some constant c. Fletcher and Rangan provided an improved scaling law on the number of measurements and also showed that the scaling law works for noisy scenarios for which the signal-to-noise (SNR) goes to infinity [12].

Another direction is to develop deterministic conditions for the exact support recovery with the OMP algorithm [13]–[21]. Those conditions are often characterized by the properties of measurement matrices, such as the mutual incoherence property (MIP) [22] and the restricted isometry property (RIP) [23]. A measurement matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a constant $c \in [0, 1)$ such that

$$(1 - c)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + c)\|\mathbf{x}\|_2^2 \tag{3}$$

for all K-sparse vectors $\mathbf{x}$. In particular, the minimum value among all constants c satisfying (3) is called the isometry constant δ_K. In the noise-free case (i.e., when the noise vector $\mathbf{v} = \mathbf{0}$), Davenport and Wakin showed that [14]

$$\delta_{K+1} < \frac{1}{3\sqrt{K}} \tag{4}$$

is sufficient for OMP to accurately recover the support of the input signal. For further improvements on this condition,

TABLE I
THE OMP ALGORITHM

Input	$\mathbf{\Phi}$, $\mathbf{y}$, and sparsity level K.
Initialize	iteration counter $k = 0$,
	estimated support $\mathcal{T}^0 = \emptyset$,
	and residual vector $\mathbf{r}^0 = \mathbf{y}$.
While	$k < K$ **do**
	$k = k + 1$.
	Identify[1] $t^k = \underset{i \in \Omega \setminus \mathcal{T}^{k-1}}{\arg\max} |\langle \phi_i, \mathbf{r}^{k-1} \rangle|$.
	Enlarge $\mathcal{T}^k = \mathcal{T}^{k-1} \cup t^k$.
	Estimate $\mathbf{x}^k = \underset{\mathbf{u}:supp(\mathbf{u})=\mathcal{T}^k}{\arg\min} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$.
	Update $\mathbf{r}^k = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^k$.
End	
Output	the estimated support $\mathcal{T}^K$ and vector $\mathbf{x}^K$.

see [15]–[20], [24], [25]. The deterministic conditions on the exact support recovery with OMP in the noisy case have been studied in [9], [25]–[28], in which the researchers considered the OMP algorithm with residual-based stopping rules and established sufficient conditions for the exact support recovery that depend on the properties of measurement matrices and the minimum magnitude of the nonzero elements of the signal.

The main purpose of this paper is to investigate deterministic conditions of OMP for the support recovery in the noisy case. Unlike previous studies that considered residual-based stopping rules for the OMP algorithm, we simply consider that OMP runs K iterations before stopping, which is arguably the most natural stopping rule if one is concerned with the recovery of exact support. We establish necessary and sufficient conditions for the exact support recovery with OMP, expressed as lower bounds on the SNR. Our results indicate that the OMP algorithm can accurately recover the support of all K-sparse signals only when the SNR scales at least linearly with the sparsity K. For high-dimensional setting, this essentially requires the SNR to be unbounded from above.

We also study the situation where the SNR is upper bounded so that the necessary condition for the exact support recovery with OMP is not fulfilled. The analysis of OMP with bounded SNR has been an interesting open problem [12]. We consider a practical case where the SNR is an absolute constant independent of the sparsity K. Our result shows that under appropriate conditions on the SNR and the isometry constant, OMP can approximately recover the support of sparse signals with only a small constant fraction of errors.

The main contributions of this paper are summarized as follows.

i) We consider the exact support recovery with OMP in the noisy scenario. Our analysis shows that OMP can accurately recover the support of any K-sparse signal if

$$\sqrt{\text{SNR}} > \frac{2\sqrt{K}(1 + \delta_{K+1})}{(1 - (\sqrt{K}+1)\delta_{K+1}) \cdot \sqrt{\text{MAR}}}, \quad (5)$$

[1]There is another popular version of OMP in the literature which normalizes columns of $\mathbf{\Phi}$ before the identification step (see, e.g., [12]). In the present paper we exclusively consider the version of OMP without column normalization, since results obtained for this version can be readily extended to the version with normalization via rescaling. Note that when $\mathbf{\Phi}$ has normalized columns (e.g., dictionary atoms), the two versions of OMP become identical.

where MAR is the minimum-to-average ratio of the input signal (see definitions of the MAR and the SNR in Section III). We also establish a necessary condition for the exact support recovery with OMP as

$$\sqrt{\text{SNR}} > \frac{\sqrt{K}(1 + \delta_{K+1})}{(1 - \sqrt{K}\delta_{K+1}) \cdot \sqrt{\text{MAR}}}. \quad (6)$$

Therefore, to ensure the perfect support recovery of all K-sparse signals with OMP, the SNR must at least scale linearly with the sparsity level K of input signals.

ii) We also consider the case where SNR is an absolute constant independent of the signal sparsity K. Our analysis shows that if

$$\sqrt{\text{SNR}} \geq \frac{\kappa}{\delta_{2K}^{3/4}} \quad (7)$$

where $\kappa := \max_{i,j \in supp(\mathbf{x})} \frac{|x_i|}{|x_j|}$, then OMP can recover the support of any K-sparse signal with error rate

$$\rho_{\text{error}} \leq C\kappa^2 \delta_{2K}^{1/2}, \quad (8)$$

where C is a constant. Therefore, with properly chosen isometry constants, the fraction of errors in the support recovery can be made arbitrarily small.

The rest of this paper is organized as follows: In Section II, we introduce notations and lemmas that are used in this paper. In Section III, we analyze necessary and sufficient conditions for the exact support recovery with OMP. In Section IV, we provide the results of OMP for the approximate support recovery of sparse signals. Concluding remarks are given in Section V.

II. PRELIMINARIES

A. Notations

We briefly summarize notations used in this paper. Let $\Omega = \{1, 2, \cdots, n\}$ and $\mathcal{T} = supp(\mathbf{x}) = \{i|i \in \Omega, x_i \neq 0\}$ denote the support of vector $\mathbf{x}$. For $\mathcal{S} \subseteq \Omega$, $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. $\mathcal{T} \setminus \mathcal{S}$ is the set of all elements contained in $\mathcal{T}$ but not in $\mathcal{S}$. $\mathbf{x}_\mathcal{S} \in \mathcal{R}^{|\mathcal{S}|}$ represents a restriction of the vector $\mathbf{x}$ to the elements with indices in $\mathcal{S}$. $\mathbf{\Phi}_\mathcal{S} \in \mathcal{R}^{m \times |\mathcal{S}|}$ is a submatrix of $\mathbf{\Phi}$ that only contains columns indexed by $\mathcal{S}$. If $\mathbf{\Phi}_\mathcal{S}$ is full column rank, then $\mathbf{\Phi}_\mathcal{S}^\dagger = (\mathbf{\Phi}_\mathcal{S}'\mathbf{\Phi}_\mathcal{S})^{-1}\mathbf{\Phi}_\mathcal{S}'$ is the pseudoinverse of $\mathbf{\Phi}_\mathcal{S}$. $span(\mathbf{\Phi}_\mathcal{S})$ represents the span of columns in $\mathbf{\Phi}_\mathcal{S}$. $\mathbf{P}_\mathcal{S} = \mathbf{\Phi}_\mathcal{S}\mathbf{\Phi}_\mathcal{S}^\dagger$ stands for the projection onto $span(\mathbf{\Phi}_\mathcal{S})$. $\mathbf{P}_\mathcal{S}^\perp = \mathbf{I} - \mathbf{P}_\mathcal{S}$ is the projection onto the orthogonal complement of $span(\mathbf{\Phi}_\mathcal{S})$, where $\mathbf{I}$ denotes the identity matrix.

B. Lemmas

The following lemmas are useful for our analysis.

Lemma 2.1 (Lemma 3 in [23]): If a measurement matrix satisfies the RIP of both orders K_1 and K_2 where $K_1 \leq K_2$, then

$$\delta_{K_1} \leq \delta_{K_2}.$$

This property is often referred to as the monotonicity of the isometry constant.

Lemma 2.2 (Direct consequences of RIP [29], [30]): Let $\mathcal{S} \subseteq \Omega$. If $\delta_{|\mathcal{S}|} < 1$ then for any $\mathbf{u} \in \mathcal{R}^{|\mathcal{S}|}$,

$$(1 - \delta_{|\mathcal{S}|}) \|\mathbf{u}\|_2 \leq \|\mathbf{\Phi}'_{\mathcal{S}} \mathbf{\Phi}_{\mathcal{S}} \mathbf{u}\|_2 \leq (1 + \delta_{|\mathcal{S}|}) \|\mathbf{u}\|_2 ,$$

$$\frac{1}{1 + \delta_{|\mathcal{S}|}} \|\mathbf{u}\|_2 \leq \|(\mathbf{\Phi}'_{\mathcal{S}} \mathbf{\Phi}_{\mathcal{S}})^{-1} \mathbf{u}\|_2 \leq \frac{1}{1 - \delta_{|\mathcal{S}|}} \|\mathbf{u}\|_2 .$$

Lemma 2.3 (Consequences of RIP [31], [32]): Let $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3 \subseteq \Omega$ and $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_2 = \emptyset$. If $\delta_{|\mathcal{S}_1|+|\mathcal{S}_2|+|\mathcal{S}_3|} < 1$, then for any vector $\mathbf{v} \in \mathcal{R}^{|\mathcal{S}_2|}$,

$$\|\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{\Phi}_{\mathcal{S}_2} \mathbf{v}\|_2 \leq \delta_{|\mathcal{S}_1|+|\mathcal{S}_2|} \|\mathbf{v}\|_2 ,$$

$$\|\mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{P}^{\perp}_{\mathcal{S}_3} \mathbf{\Phi}_{\mathcal{S}_2} \mathbf{v}\|_2 \leq \delta_{|\mathcal{S}_1|+|\mathcal{S}_2|+|\mathcal{S}_3|} \|\mathbf{v}\|_2 .$$

Lemma 2.4 (Proposition 3.1 in [29]): Let $\mathcal{S} \subseteq \Omega$. If $\delta_{|\mathcal{S}|} < 1$, then for any vector $\mathbf{u} \in \mathcal{R}^m$,

$$\|\mathbf{\Phi}'_{\mathcal{S}} \mathbf{u}\|_2 \leq \sqrt{1 + \delta_{|\mathcal{S}|}} \|\mathbf{u}\|_2.$$

Lemma 2.5 (Lemma 5 in [26]): Let $\mathcal{S}_1, \mathcal{S}_2, \subseteq \Omega$ be disjoint sets and denote $\mathbf{A} = \mathbf{\Phi}'_{\mathcal{S}_1} \mathbf{P}^{\perp}_{\mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1}$ and $\mathbf{B} = \mathbf{\Phi}'_{\mathcal{S}_1 \cup \mathcal{S}_2} \mathbf{\Phi}_{\mathcal{S}_1 \cup \mathcal{S}_2}$. Then the minimum and maximum eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ satisfy

$$\lambda_{\min}(\mathbf{A}) \geq \lambda_{\min}(\mathbf{B}) \quad \text{and} \quad \lambda_{\max}(\mathbf{A}) \leq \lambda_{\max}(\mathbf{B}).$$

III. EXACT SUPPORT RECOVERY VIA OMP IN THE PRESENCE OF NOISE

A. Main Results

In this section, we analyze the condition for the exact support recovery with OMP in the presence of noise. Following the analysis in [12], we parameterize the dependence on the noise $\mathbf{v}$ and the signal $\mathbf{x}$ with two quantities:

$$\text{SNR} := \frac{\|\mathbf{\Phi}\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2}.$$

and

$$\text{MAR} = \frac{\min_{j \in \mathcal{T}} |x_j|^2}{\|\mathbf{x}\|_2^2 / K}.$$

The next theorem provides a condition under which OMP can accurately recover the support of any K-sparse signal $\mathbf{x}$.

Theorem 3.1 (Sufficient Condition): Suppose that the measurement matrix $\mathbf{\Phi}$ satisfies the RIP with $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Then OMP performs the exact support recovery of any K-sparse signal $\mathbf{x}$ from its noisy measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$, provided that

$$\sqrt{\text{SNR}} > \frac{2\sqrt{K}(1 + \delta_{K+1})}{(1 - (\sqrt{K}+1)\delta_{K+1}) \cdot \sqrt{\text{MAR}}}. \tag{9}$$

One can interpret from Theorem 3.1 that in the high-dimensional setting, the exact support recovery with OMP can be ensured in the high SNR region. Moreover, observe that (9) can be rewritten as

$$\delta_{K+1} < \frac{1 - t}{\sqrt{K} + 1 + t}, \tag{10}$$

where

$$t := \frac{2\sqrt{K}}{\sqrt{\text{SNR} \cdot \text{MAR}}}.$$

Hence, when $t \to 0$, the condition reduces to

$$\delta_{K+1} < \frac{1}{\sqrt{K}+1}, \tag{11}$$

which coincides with the recovery condition of OMP in the noise-free case [19], [20]. The condition in (11) has also been shown to be nearly necessary for the exact support recovery with OMP since it cannot be further relaxed to $\delta_{K+1} \leq \frac{1}{\sqrt{K}}$ [19], [20], [33].

The following theorem gives a necessary condition for the exact support recovery with the OMP algorithm.

Theorem 3.2 (Necessary Condition): If one wish to accurately recover the support of any K-sparse signal $\mathbf{x}$ from its noisy measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$ with OMP, then the SNR should satisfy

$$\sqrt{\text{SNR}} > \frac{\sqrt{K}(1 + \delta_{K+1})}{(1 - \sqrt{K}\delta_{K+1}) \cdot \sqrt{\text{MAR}}}. \tag{12}$$

Proof: See Appendix A. ∎

Remark 1:

i) Loosely speaking, the lower bounds in (9) and (12) can be matched within a constant factor of two. Interestingly, since MAR ≤ 1 and $\delta_{K+1} \geq 0$, one can directly obtain from (12) that

$$\text{SNR} > K, \tag{13}$$

which implies that in order for OMP to accurately recover the support of any K-sparse signal, the SNR must at least scale linearly with the sparsity level K of the signal. For high-dimensional setting, this essentially requires the SNR to be unbounded from above.

ii) The sufficient condition for the exact support recovery with OMP under a similar SNR setting has been studied in [12], in which the authors considered large random measurement matrices. They provided asymptotic and probabilistic results on the scaling law for the number of measurements that ensures the exact support recovery of every fixed K-sparse signal when the SNR approaches to infinity. Our result in Theorem 3.1 extends that in [12] by providing a deterministic condition that applies to general measurement matrices and also holds uniformly for all K-sparse signals.

iii) The result in Theorem 3.1 is also closely related to the results in [9], [25]–[28], in which the authors considered the OMP algorithm with residual-based stopping rules by assuming that the noise level is known a priori. They established conditions that depend on the properties of measurement matrices and the minimum magnitude of nonzero elements in the signal $\mathbf{x}$. In fact, it can be shown that the conditions proposed in [9], [25]–[28] essentially impose a similar requirement on the size of measurements as our result established in Theorem 3.1. However, our analysis differs in 1) that we characterize the sufficient

condition for the exact support recovery with a lower bound on the SNR and 2) that we consider the stopping rule that the OMP algorithm runs exact K iterations before stopping and does not require the assumption of knowing the noise level. In addition, we provide in Theorem 3.2 the necessary condition analysis for the exact support recovery with OMP, for which there is no counterpart in the studies of [9], [25]–[28].

iv) For the OMP algorithm, it had become usual, when recovering a K-sparse signal, to consider the performance of the algorithm after K iterations. See, for instance, [11]–[15], [19], [20], [25]. Compared to stopping rules that are based on the residual tolerance (e.g., [26]) or the number of maximally allowed iterations (e.g., $\lceil cK \rceil$ iterations, $c > 1$ [17], [18], [34]–[36]), running OMP for K iterations is very natural in that it directly allows the algorithm to recover the exact support set without false alarms or missed detections since the K-th iterate is itself a K-sparse signal. This feature is appealing when one is concerned with the exact support recovery.

v) It is worth mentioning that for the noise-free case, the stopping rule of OMP does not require additional information about the data, since it can simply be $\|\mathbf{r}^k\|_2 = 0$ (or $\|\mathbf{\Phi}'\mathbf{r}^{k-1}\|_\infty = 0$).[2] Whereas, in the noisy case, some prior information on the signal and the noise is often needed. For example, residual-based stopping rules may rely on the prior knowledge of the noise level [9], [25]–[28], and running OMP for K iterations requires to know a priori the sparsity level K of the input signal. The assumption of knowing the sparsity (or its region) has been commonly made for the algorithm design and the performance analysis in the CS field (see, e.g., [29], [37]–[41]). However, it should be noted that in many applications, the sparsity K is often not available and moreover, the underlying signal may not be exactly sparse. A typical scenario is that the input signal is not exactly sparse but only approximately sparse with a few significant nonzero coefficients.[3] For this scenario, our analysis for the recovery of the exactly sparse signal can be readily extended to the recovery of significant nonzero coefficients of the signal, by treating the contribution of those small nonzero coefficients as part of noise and utilizing the techniques developed in [26], [41].

B. Proof of Theorem 3.1

Our proof of Theorem 3.1 is essentially an extension of the proof technique in [42, Theorem 3.4 and 3.5]. Note that [42, Theorem 3.4 and 3.5] studied the recovery condition for the generalized OMP (gOMP) algorithm in the noise-free situation. Our contribution is to generalize the analysis to the noisy case. We would like to mention that [42] also

[2]Without noise, the stopping rule of $\|\mathbf{r}^k\|_2 = 0$ (or $\|\mathbf{\Phi}'\mathbf{r}^{k-1}\|_\infty = 0$) is essentially equivalent to running OMP for K iterations.

[3]Note that if the signal to be recovered is not even approximately sparse, then compressed sensing technique may not apply.

provided a noisy case analysis but focused only on the ℓ_2-norm distortion of the signal recovery and the corresponding result is also weaker than the result established in this paper (see Section III-C).

The proof works by mathematical induction. For the convenience of stating the results, we say that OMP makes a success at an iteration if it selects a correct index at the iteration. We will first give a condition that guarantees the success of OMP at the first iteration. Then we will assume that OMP has been successful in the previous k ($1 \le k < K$) iterations and will derive a condition under which OMP also makes a success at the $(k+1)$-th iteration. Finally, we will combine the two conditions to establish an overall condition for the OMP algorithm.

- **Success at the first iteration**:

  From Table I, we know that at the first iteration, OMP selects the index t^1 corresponding to the column $\phi_{t^1} \in \mathbf{\Phi}$ that is most strongly correlated with the measurement vector $\mathbf{y}$. Hence,

$$
\begin{aligned}
|\langle \phi_{t^1}, \mathbf{y} \rangle| &= \max_{i \in \Omega} |\langle \phi_i, \mathbf{y} \rangle| \\
&\ge \max_{i \in \mathcal{T}} |\langle \phi_i, \mathbf{y} \rangle| \\
&\ge \sqrt{\frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} \langle \phi_i, \mathbf{y} \rangle^2} \\
&= \frac{1}{\sqrt{K}} \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{y}\|_2 \\
&\overset{(a)}{=} \frac{1}{\sqrt{K}} \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}} + \mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \\
&\overset{(b)}{\ge} \frac{1}{\sqrt{K}} \left(\|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}}\|_2 - \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \right) \\
&\overset{(c)}{\ge} \frac{1}{\sqrt{K}} \left((1 - \delta_K) \|\mathbf{x}\|_2 - \|\mathbf{\Phi}'_{\mathcal{T}} \mathbf{v}\|_2 \right) \\
&\overset{(d)}{\ge} \frac{1}{\sqrt{K}} \left((1 - \delta_K) \|\mathbf{x}\|_2 - \sqrt{1 + \delta_K}\, \|\mathbf{v}\|_2 \right),
\end{aligned}
\tag{14}
$$

  where (a) is because $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$, (b) is from the triangle inequality, (c) is from the RIP, and (d) is due to Lemma 2.4.

  On the other hand, if a wrong index is chosen at the first iteration (i.e., $t^1 \notin \mathcal{T}$), then

$$
\begin{aligned}
|\langle \phi_{t^1}, \mathbf{y} \rangle| &= |\langle \phi_{t^1}, \mathbf{\Phi}\mathbf{x} \rangle + \langle \phi_{t^1}, \mathbf{v} \rangle| \\
&\overset{(a)}{\le} |\langle \phi_{t^1}, \mathbf{\Phi}\mathbf{x} \rangle| + |\langle \phi_{t^1}, \mathbf{v} \rangle| \\
&= \|\phi'_{t^1} \mathbf{\Phi}_{\mathcal{T}} \mathbf{x}_{\mathcal{T}}\|_2 + \|\phi'_{t^1} \mathbf{v}\|_2 \\
&\overset{(b)}{\le} \delta_{K+1} \|\mathbf{x}\|_2 + \|\phi'_{t^1} \mathbf{v}\|_2 \\
&\overset{(c)}{\le} \delta_{K+1} \|\mathbf{x}\|_2 + \sqrt{1 + \delta_1} \|\mathbf{v}\|_2, \tag{15}
\end{aligned}
$$

  where (a) uses the triangle inequality, (b) follows from Lemma 2.3 and (c) is due to Lemma 2.4.

This, however, contradicts (14) whenever

$$\delta_{K+1}\|\mathbf{x}\|_2 + \sqrt{1+\delta_1}\ \|\mathbf{v}\|_2$$
$$< \frac{1}{\sqrt{K}}\left((1-\delta_K)\|\mathbf{x}\|_2 - \sqrt{1+\delta_K}\ \|\mathbf{v}\|_2\right). \quad (16)$$

Since $\delta_1 \le \delta_K \le \delta_{K+1}$ (by the monotonicity of isometry constant), (16) is guaranteed by

$$\delta_{K+1}\|\mathbf{x}\|_2 + \sqrt{1+\delta_{K+1}}\|\mathbf{v}\|_2$$
$$< \frac{1}{\sqrt{K}}\left((1-\delta_{K+1})\|\mathbf{x}\|_2 - \sqrt{1+\delta_{K+1}}\|\mathbf{v}\|_2\right). \quad (17)$$

Or equivalently,

$$\left(\frac{1-\delta_{K+1}}{\sqrt{K}} - \delta_{K+1}\right)\|\mathbf{x}\|_2$$
$$> \left(1 + \frac{1}{\sqrt{K}}\right)\sqrt{1+\delta_{K+1}}\|\mathbf{v}\|_2. \quad (18)$$

Furthermore, since

$$\|\mathbf{\Phi}\mathbf{x}\|_2 \le \sqrt{1+\delta_K}\|\mathbf{x}\|_2 \le \sqrt{1+\delta_{K+1}}\|\mathbf{x}\|_2, \quad (19)$$

and also noting that $\text{SNR} = \frac{\|\mathbf{\Phi}\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2}$, we can show that (18) holds true if

$$\frac{1-\delta_{K+1}}{\sqrt{K}} - \delta_{K+1} > \frac{1+\delta_{K+1}}{\sqrt{\text{SNR}}}\left(1 + \frac{1}{\sqrt{K}}\right). \quad (20)$$

That is,

$$\sqrt{\text{SNR}} > \frac{(\sqrt{K}+1)(1+\delta_{K+1})}{1-(\sqrt{K}+1)\delta_{K+1}}, \quad (21)$$

Therefore, under (21), a correct index is chosen at the first iteration of OMP.

- **Success at the general iteration**:
  Assume that OMP has been successful in each of the previous k ($1 \le k < K$) iterations. Then,

  $$|\mathcal{T} \cap \mathcal{T}^k| = k. \quad (22)$$

  Under this assumption, we will derive a condition that ensures OMP to make a success at the $(k+1)$-th iteration as well.

  For analytical convenience, we introduce two quantities. Let u denote the largest value in $\{|\langle \phi_i, \mathbf{r}^k \rangle|\}_{i \in \mathcal{T} \setminus \mathcal{T}^k}$ and let v denote the largest value in $\{|\langle \phi_i, \mathbf{r}^k \rangle|\}_{i \in \Omega \setminus (\mathcal{T} \cup \mathcal{T}^k)}$. Note that $\mathcal{T} \setminus \mathcal{T}^k$ and $\Omega \setminus (\mathcal{T} \cup \mathcal{T}^k)$ are the set of remaining correct indices and the set of remaining incorrect indices, respectively. Then it is clear that if

  $$u > v, \quad (23)$$

  a good index will be selected at the $(k+1)$-th iteration of OMP. The following proposition characterizes the lower bound of u and the upper bound of v.

  Proposition 3.3:

  $$u \ge \frac{1}{\sqrt{K-k}}\left((1-\delta_K)\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 - \sqrt{1+\delta_K}\|\mathbf{v}\|_2\right),$$

  $$v \le \delta_{K+1}\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 + \sqrt{1+\delta_1}\ \|\mathbf{v}\|_2.$$

  Proof: See Appendix B. ∎

Using Proposition 3.3, we obtain the sufficient condition to $u > v$ as

$$\frac{1}{\sqrt{K-k}}\left((1-\delta_K)\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 - \sqrt{1+\delta_K}\ \|\mathbf{v}\|_2\right)$$
$$> \delta_{K+1}\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2 + \sqrt{1+\delta_1}\ \|\mathbf{v}\|_2,$$

or equivalently,

$$\left(\frac{1-\delta_{K+1}}{\sqrt{K-k}} - \delta_{K+1}\right)\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|_2$$
$$> \left(1 + \frac{1}{\sqrt{K-k}}\right)\sqrt{1+\delta_{K+1}}\|\mathbf{v}\|_2. \quad (24)$$

Note that

$$\|\mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^k}\|$$
$$\ge \sqrt{|\mathcal{T}\setminus\mathcal{T}^k|}\min_{j\in\mathcal{T}}|x_j|$$
$$\overset{(a)}{\ge} \sqrt{K-k}\left(\frac{\sqrt{\text{MAR}}\cdot\|\mathbf{x}\|_2}{\sqrt{K}}\right)$$
$$\overset{(b)}{\ge} \sqrt{\frac{K-k}{K(1+\delta_{K+1})}}\cdot\sqrt{\text{MAR}}\cdot\|\mathbf{\Phi}\mathbf{x}\|_2$$
$$\overset{(c)}{=} \sqrt{\frac{K-k}{K(1+\delta_{K+1})}}\cdot\sqrt{\text{MAR}\cdot\text{SNR}}\cdot\|\mathbf{v}\|_2, \quad (25)$$

where (a) is from the definition of MAR, (b) is from (19), and (c) is because $\text{SNR} = \frac{\|\mathbf{\Phi}\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2}$.

Using (24) and (25), we can show $u > v$ holds true if

$$\left(\frac{1-\delta_{K+1}}{\sqrt{K-k}} - \delta_{K+1}\right)\sqrt{\frac{K-k}{K(1+\delta_{K+1})}}\cdot\sqrt{\text{MAR}\cdot\text{SNR}}$$
$$> \left(1 + \frac{1}{\sqrt{K-k}}\right)\sqrt{1+\delta_{K+1}}. \quad (26)$$

That is,

$$\sqrt{\text{SNR}} > \frac{(1+\delta_{K+1})(\sqrt{K-k}+1)\sqrt{K}}{(1-(\sqrt{K-k}+1)\delta_{K+1})\sqrt{K-k}\cdot\sqrt{\text{MAR}}}. \quad (27)$$

Furthermore, observe that

$$\frac{(1+\delta_{K+1})(\sqrt{K-k}+1)\sqrt{K}}{(1-(\sqrt{K-k}+1)\delta_{K+1})\sqrt{K-k}\cdot\sqrt{\text{MAR}}}$$
$$= \frac{\sqrt{K}(1+\delta_{K+1})}{(1-(\sqrt{K-k}+1)\delta_{K+1})\cdot\sqrt{\text{MAR}}}$$
$$\times \left(1 + \frac{1}{\sqrt{K-k}}\right)$$
$$\overset{(a)}{<} \frac{2\sqrt{K}(1+\delta_{K+1})}{(1-(\sqrt{K}+1)\delta_{K+1})\cdot\sqrt{\text{MAR}}}, \quad (28)$$

where (a) is from the assumption that $1 \le k < K$ and hence $1 \le \sqrt{K-k} < \sqrt{K}$.

Hence, using (27) and (28), we can show that $u > v$ is also ensured by

$$\sqrt{\text{SNR}} > \frac{2\sqrt{K}(1+\delta_{K+1})}{(1-(\sqrt{K}+1)\delta_{K+1})\cdot\sqrt{\text{MAR}}}, \quad (29)$$

Therefore, under (29), OMP makes a success at the $(k+1)$-th iteration.

So far, we have obtained condition (21) for the success of the first iteration and condition (29) for the success of the general iteration. We now combing the two conditions to obtain an overall condition that ensures the selection of all support indices with the OMP algorithm.

Clearly the overall condition will be determined by the more restrictive one between conditions (21) and (29). Thus we compare the right-hand-side of (21) and (29). Since

$$\frac{2(1+\delta_{K+1})\sqrt{K}}{(1-(\sqrt{K}+1)\delta_{K+1})\cdot\sqrt{\mathrm{MAR}}} \geq \frac{(1+\delta_{K+1})(\sqrt{K}+1)}{1-(\sqrt{K}+1)\delta_{K+1}},$$

condition (29) is more restrictive than (21) and hence becomes the overall condition for the OMP algorithm. The proof is thus complete.

C. Recovery Distortion in ℓ_2-norm

When all support indices of $\mathbf{x}$ have been recovered with OMP (i.e., $\mathcal{T}^K = \mathcal{T}$), we have

$$
\begin{aligned}
\|\mathbf{x}-\mathbf{x}^K\|_2 &\overset{(a)}{=} \|\mathbf{x}_{\mathcal{T}^K}-\boldsymbol{\Phi}_{\mathcal{T}^K}^\dagger\mathbf{y}\|_2 \\
&= \|\mathbf{x}_{\mathcal{T}^K}-\boldsymbol{\Phi}_{\mathcal{T}^K}^\dagger(\boldsymbol{\Phi}_{\mathcal{T}}\mathbf{x}_{\mathcal{T}}+\mathbf{v})\|_2 \\
&\overset{(b)}{=} \|\mathbf{x}_{\mathcal{T}^K}-\boldsymbol{\Phi}_{\mathcal{T}^K}^\dagger(\boldsymbol{\Phi}_{\mathcal{T}^K}\mathbf{x}_{\mathcal{T}^K}+\mathbf{v})\|_2 \\
&= \|\boldsymbol{\Phi}_{\mathcal{T}^K}^\dagger\mathbf{v}\|_2 \\
&\overset{(c)}{\leq} \frac{\|\boldsymbol{\Phi}_{\mathcal{T}^K}\boldsymbol{\Phi}_{\mathcal{T}^K}^\dagger\mathbf{v}\|_2}{\sqrt{1-\delta_{|\mathcal{T}^K|}}} \\
&= \frac{\|\mathbf{P}_{\mathcal{T}^K}\mathbf{v}\|_2}{\sqrt{1-\delta_K}} \\
&\leq \frac{\|\mathbf{v}\|_2}{\sqrt{1-\delta_K}},
\end{aligned}
\tag{30}
$$

where (a) is because $\mathcal{T}^K = \mathcal{T}$ and

$$\mathbf{x}^K = \underset{\mathbf{u}:supp(\mathbf{u})=\mathcal{T}^K}{\arg\min} \|\mathbf{y}-\boldsymbol{\Phi}\mathbf{u}\|_2,$$

(b) is due to $\mathcal{T}^K = \mathcal{T}$, and (c) is from the RIP.

One can interpret from (30) that the upper bound of the ℓ_2-norm recovery distortion with OMP is just proportional the noise energy, which outperforms the result in [42] that suggested a recovery distortion upper bounded by $\mathcal{O}(\sqrt{K})\|\mathbf{v}\|_2$.

IV. APPROXIMATE SUPPORT RECOVERY VIA OMP IN THE PRESENCE OF NOISE

In the last section we have shown that the exact support recovery with OMP requires the SNR to scale linearly with the sparsity lever K. For high-dimensional setting, this would require the SNR to be unbounded. However, in practical applications, we are often facing with the situation where the SNR is bounded from above. A particularly interesting case might be the case where the SNR is an absolute constant independent of the problem size. In this case, of course, the necessary condition (in Section III) is not fulfilled so that the exact support recovery of all sparse signals with OMP

is impossible. However, we will show that recovery with an arbitrarily small but constant fraction of errors is possible.

The following theorem demonstrates that for properly chosen isometry constants, there exists an absolute constant SNR, under which OMP can approximately recover the support of any K-sparse signal with a small constant fraction of errors.

Theorem 4.1: Let $\kappa := \max_{i,j\in\mathcal{T}} \frac{|x_i|}{|x_j|}$. Then if SNR $\geq \kappa^2\delta_{2K}^{-3/2}$, OMP recovers the support of K-sparse signal $\mathbf{x}$ from its noisy measurements $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{v}$ with error rate

$$\rho_{\mathrm{error}} \leq C\kappa^2\delta_{2K}^{1/2}$$

where C is a constant.

Remark 2:
i) It is intuitively easy to see that the bound of error rate in Theorem 4.1 is reasonable because in the special case of orthonormal matrix $\boldsymbol{\Phi}$ (i.e., $\delta_{2K} = 0$), the result in Theorem 4.1 suggests that if SNR $= \infty$, then the error rate $\rho = 0$, which matches with the trivial fact that when there is no noise and $\boldsymbol{\Phi}$ is an orthonormal matrix, OMP can identify a correct index at each iteration and will accurately recover the whole support of signal $\mathbf{x}$ in exact K iterations.

ii) An interesting point we would like to mention is that our result for the approximate support recovery with OMP only requires the isometry constant δ_{2K} to be an absolute constant, which essentially imposes a mild constraint on the measurement matrix $\boldsymbol{\Phi}$. For example, for random Gaussian measurement matrices, it can be satisfied with

$$m \geq cK\log\frac{n}{K} \tag{31}$$

for some constant c [23], [43]. In CS, this implies that OMP can essentially perform the approximate support recovery of sparse signals with optimal number of random measurements up to a constant.

It is well known that recovering sparse signals with nonzero elements of same magnitude is a particularly challenging case for the OMP algorithm [17], [39]. The following corollary provides the result of OMP on the approximate support recovery for this type of input signals.

Corollary 4.2: Consider K-sparse signals $\mathbf{x}$ with nonzero elements of equal magnitude. Then if SNR $\geq \delta_{2K}^{-3/2}$, OMP can recover the support of $\mathbf{x}$ from its noisy measurements $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{v}$ with error rate

$$\rho_{\mathrm{error}} \leq C\delta_{2K}^{1/2}$$

where C is a constant.

Our analysis is inspired by the recent work of Livshitz and Temlyakov [36] and will also rely on some proof techniques in [35]. The main idea behind our analysis is that most of support indices can essentially be identified in K iterations of OMP and the number of false alarms is small. The proof of Theorem 4.1 follows along a similar line as the proofs in [36], but with three important distinctions. Firstly, the main goals of proofs are different. While our analysis is based on

the approximate support recovery of K-sparse signals with K iterations of OMP, the analysis in [36] concerned the exact support recovery with OMP in more than K iterations. Secondly, compared to the result in [36], our result is more general in that it applies to input signals with arbitrary sparsity level K and with nonzero elements of arbitrary magnitudes. Note that the analysis of [36] assumed that $K \geq \delta_{2K}^{-1/2}$, which essentially applies to the situation where the sparsity K of the input signal is nontrivial. In addition, [36] considered only the recovery of signals with magnitudes of nonzero elements upper bounded by one. Thirdly, and most importantly, we consider the scenario where the measurement noise is present and build conditions based on the SNR. Whereas, the analyses in [36] focused only on the situation without noise.

A. Proof of Theorem 4.1

Before we proceed to the details of the proof, we introduce some useful notations and definitions. For notational simplicity, let $\delta := \delta_{2K}$. At the k-th iteration ($0 \leq k \leq K$), let $\Gamma^k := \mathcal{T} \setminus \mathcal{T}^k$ denote the set of missed detection of support indices. For given constant $\tau \in (0, 1]$, let Γ_τ^k denote the subset of Γ^k corresponding to the $\lceil \tau K \rceil$ largest elements (in magnitude) of $\mathbf{x}_{\Gamma^k}$. Also, let x_τ^k denote the $\lceil \tau K \rceil$-th largest element (in magnitude) in $\mathbf{x}_{\Gamma^k}$. Following the idea in [36], we will fix $\tau = \delta^{1/2}$. If $\lceil \tau K \rceil > |\Gamma^k|$, then set $\Gamma_\tau^k = \Gamma^k$ and $x_\tau^k = 0$. Since OMP totally runs K iterations before stopping, the error rate of the support recovery can be given by

$$\rho_{\text{error}} := \frac{|\mathcal{T}^K \setminus \mathcal{T}|}{|\mathcal{T}|}. \tag{32}$$

The proof of Theorem 4.1 consists of two parts. In the first part, we will provide a lower bound on the reduction of residual energy at each iteration of OMP (Proposition 4.3). In the second part, by means of the lower bound obtained in Proposition 4.3, we will estimate the remaining energy in the residual vector $\mathbf{r}^K$. The estimate of the energy of $\mathbf{r}^K$ will then allow us to derive an upper bound on the number of missed detections (i.e., $|\Gamma^K|$). Since the OMP algorithm totally chooses K indices, it is easy to see that the number missed detections (after K iterations) is equal to the number of false alarms. i.e.,

$$|\Gamma^K| = |\mathcal{T}^K \setminus \mathcal{T}|, \tag{33}$$

which implies that

$$\rho_{\text{error}} = \frac{|\Gamma^K|}{K}. \tag{34}$$

Therefore, from the upper bound of $|\Gamma^K|$ we can directly obtain an upper bound of the error rate for the support recovery with OMP.

Proposition 4.3: For any $0 \leq k \leq K - \lceil \delta^{1/2} K \rceil$, the residual of OMP satisfies

$$\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 \geq (1 - 7\delta^{1/2}) \left(x_{\delta^{1/2}}^k\right)^2.$$

Proof: We shall prove Proposition 4.3 in two steps. First, we show that the residual power difference of OMP satisfies

(see Appendix C)[4]

$$\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 \geq \frac{\|\mathbf{\Phi}'\mathbf{r}^k\|_\infty^2}{1 + \delta_1}. \tag{35}$$

In the second step, we show that (see Appendix D)

$$\|\mathbf{\Phi}'\mathbf{r}^k\|_\infty^2 \geq (1 - 7\tau) \left(x_\tau^k\right)^2. \tag{36}$$

Using (35) and (36), we have

$$\begin{aligned}
\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 &\geq \left(\frac{1 - 6\tau}{1 + \delta_1}\right)(x_\tau^k)^2 \\
&\overset{(a)}{\geq} \left(\frac{1 - 6\tau}{1 + \delta}\right)(x_\tau^k)^2 \\
&= \left(\frac{1 - 6\tau}{1 + \tau^2}\right)(x_\tau^k)^2 \\
&= \left(1 - 7\tau + \frac{\tau - \tau^2 + 7\tau^3}{1 + \tau^2}\right)(x_\tau^k)^2 \\
&\overset{(b)}{\geq} (1 - 7\tau)(x_\tau^k)^2 \\
&= (1 - 7\delta^{1/2})\left(x_{\delta^{1/2}}^k\right)^2,
\end{aligned} \tag{37}$$

where (a) is because $\delta = \delta_{2K} \geq \delta_1$ and (b) is from $\tau = \delta^{1/2} \in [0, 1)$, which establishes the proposition. ∎

In Proposition 4.3, we have shown that each iteration of OMP makes non-trivial progress by providing the lower bound on the reduction of residual energy at each iteration. Next, using the bound obtained in Proposition 4.3, we will derive an upper bound on the number of missed detections in the support recovery of OMP. Without loss of generality we assume that $\mathcal{T} = \{1, \cdots, K\}$ and that the elements of $\{x_i\}_{i=1}^K$ are in a descending order of their magnitudes. Then from the definition of x_τ^k we have that for any $k \geq 0$, $k + \lceil \tau K \rceil \leq K$,

$$|x_\tau^k| \geq |x_{k+\lceil \tau K \rceil}|. \tag{38}$$

By applying Proposition 4.3, we have

$$\begin{aligned}
\|\mathbf{r}^K\|_2^2 &= \|\mathbf{r}^0\|_2^2 - \sum_{k=0}^{K-1} (\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2) \\
&\overset{(a)}{\leq} \|\mathbf{y}\|_2^2 - \sum_{k=1}^{K-\lceil \tau K \rceil} (\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2) \\
&\leq \|\mathbf{y}\|_2^2 - \sum_{k=1}^{K-\lceil \tau K \rceil} (1 - 7\tau)(x_\tau^k)^2 \\
&\overset{(b)}{\leq} \|\mathbf{y}\|_2^2 - \sum_{k=1}^{K-\lceil \tau K \rceil} (1 - 7\tau)(x_{k+\lceil \tau K \rceil})^2 \\
&= \|\mathbf{y}\|_2^2 - \sum_{k=\lceil \tau K \rceil + 1}^{K} (1 - 7\tau)(x_i)^2 \tag{39}
\end{aligned}$$

where (a) uses the facts that $\lceil \tau K \rceil \geq 1$ and that the energy of residual of the OMP algorithm is always non-increasing with the number of iterations (i.e., $\|\mathbf{r}^k\|_2^2 \geq \|\mathbf{r}^{k+1}\|_2^2$, $k \geq 0$), and (b) is from (38).

---

[4]This proof is essentially identical to a result in [35]. Since it will play a key role in the proof of (36), we include the proof for completeness.

Note that

$$
\begin{aligned}
\|\mathbf{y}\|_2^2 &= \|\mathbf{\Phi x} + \mathbf{v}\|_2^2 \\
&\overset{(a)}{\leq} (1+\tau)\|\mathbf{\Phi x}\|_2^2 + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{\leq} (1+\tau)(1-\delta)\|\mathbf{x}\|_2^2 + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&\overset{(c)}{\leq} (1+3\tau)\|\mathbf{x}\|_2^2 + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&= (1+3\tau)\sum_{i=1}^{K}(x_i)^2 + (1+1/\tau)\|\mathbf{v}\|_2^2, \quad (40)
\end{aligned}
$$

where (a) is from the fact that

$$
\|\mathbf{u} + \mathbf{v}\|_2^2 \leq (1+\tau)\|\mathbf{u}\|_2^2 + (1+1/\tau)\|\mathbf{v}\|_2^2
$$

with $\mathbf{u} = \mathbf{\Phi x}$, (b) is due to the RIP, and (c) is from that

$$
(1+\tau)(1+\delta) \leq (1+\tau)^2 \leq 1 + 3\tau.
$$

Hence, we can rewrite (39) as

$$
\begin{aligned}
\|\mathbf{r}^K\|_2^2 &\leq \|\mathbf{y}\|_2^2 - \sum_{k=\lceil\tau K\rceil+1}^{K}(1-7\tau)(x_i)^2 \\
&\leq (1+3\tau)\sum_{i=1}^{K}(x_i)^2 - \sum_{i=\lceil\tau K\rceil+1}^{K}(1-7\tau)(x_i)^2 \\
&\quad + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&\leq 10\tau\sum_{i=1}^{K}(x_i)^2 + \sum_{i=1}^{\lceil\tau K\rceil}(x_i)^2 + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&\leq (10\tau K + \lceil\tau K\rceil)(x_{\max})^2 + (1+1/\tau)\|\mathbf{v}\|_2^2 \\
&\leq 11\lceil\tau K\rceil(x_{\max})^2 + (1+1/\tau)\|\mathbf{v}\|_2^2. \quad (41)
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\|\mathbf{r}^K\|_2^2 &= \|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}^K) + \mathbf{v}\|_2^2 \\
&\overset{(a)}{\geq} (1-\tau)\|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}^K)\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{\geq} (1-\tau)(1-\delta)\|\mathbf{x} - \mathbf{x}^K\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2 \\
&\overset{(c)}{\geq} (1-2\tau)\|\mathbf{x} - \mathbf{x}^K\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2 \\
&\geq (1-2\tau)\|(\mathbf{x} - \mathbf{x}^K)_{\Gamma^K}\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2 \\
&\overset{(d)}{\geq} (1-2\tau)\|\mathbf{x}_{\Gamma^K}\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2, \quad (42)
\end{aligned}
$$

where (a) uses the fact that

$$
\|\mathbf{u} + \mathbf{v}\|_2^2 \geq (1-\tau)\|\mathbf{u}\|_2^2 - (1/\tau - 1)\|\mathbf{v}\|_2^2
$$

with $\mathbf{u} = \mathbf{\Phi}(\mathbf{x} - \mathbf{x}^K)$, (b) follows from the RIP, (c) is because

$$
(1-\tau)(1-\delta) \geq (1-\tau)^2 = 1 - 2\tau + \tau^2 \geq 1 - 2\tau,
$$

and (d) is due to the fact that $\mathbf{x}^K$ is supported on $\mathcal{T}^K$ and hence $\mathbf{x}_{\Gamma^K}^K = \mathbf{x}_{\mathcal{T}\setminus\mathcal{T}^K}^K = \mathbf{0}$.

Using (41) and (42), we have

$$
\begin{aligned}
\|\mathbf{x}_{\Gamma^K}\|_2^2 & \\
&\leq \frac{1}{1-2\tau}\left(11\lceil\tau K\rceil(x_{\max})^2 + \frac{2}{\tau}\|\mathbf{v}\|_2^2\right) \\
&\overset{(a)}{\leq} \frac{1}{1-2\tau}\left(11\lceil\tau K\rceil(x_{\max})^2 + 2\tau^2(1+\tau^2)K(x_{\min})^2\right) \\
&\leq \frac{1}{1-2\tau}\left(11\lceil\tau K\rceil\kappa^2 + 2\tau^2(1+\tau^2)K\right)(x_{\min})^2 \\
&\leq \left(\frac{11\kappa^2 + 2\tau(1+\tau^2)}{1-2\tau}\right)\lceil\tau K\rceil(x_{\min})^2 \\
&\leq C\kappa^2\tau K(x_{\min})^2 \\
&= C\kappa^2\delta^{1/2}K(x_{\min})^2, \quad (43)
\end{aligned}
$$

where (a) is due to the facts 1) that

$$
\text{SNR} = \frac{\|\mathbf{\Phi x}\|_2^2}{\|\mathbf{v}\|_2^2} \geq \kappa^2\delta_{2K}^{-3/2} = \frac{\kappa^2}{\tau^3},
$$

and 2) that

$$
\begin{aligned}
\|\mathbf{\Phi x}\|_2^2 &\overset{\text{RIP}}{\leq} (1+\delta)\|\mathbf{x}\|_2^2 \\
&\leq (1+\delta)K(x_{\max})^2 \\
&\leq (1+\delta)K(\kappa x_{\min})^2,
\end{aligned}
$$

and hence

$$
\begin{aligned}
\frac{2}{\tau}\|\mathbf{v}\|_2^2 &\leq 2\tau^2(1+\delta)K(x_{\min})^2 \\
&= 2\tau^2(1+\tau^2)K(x_{\min})^2.
\end{aligned}
$$

Finally, by noting that

$$
\|\mathbf{x}_{\Gamma^K}\|_2^2 \geq |\Gamma^K|(x_{\min})^2,
$$

we have that the number of missed detections satisfies

$$
|\Gamma^K| \leq C\kappa^2\delta^{1/2}K. \quad (44)
$$

Recall that the number missed detections (after K iterations) is essentially equal to the number of false alarms (i.e., $|\Gamma^K| = |\mathcal{T}^K\setminus\mathcal{T}|$). Therefore, the error rate of support recovery with OMP satisfies

$$
\rho_{\text{error}} = \frac{|\mathcal{T}^K\setminus\mathcal{T}|}{|\mathcal{T}|} = \frac{|\Gamma^K|}{K} \leq C\kappa^2\delta^{1/2}. \quad (45)
$$

The proof is now complete.

V. CONCLUSION

In this paper, we have studied the performance of OMP for the support recovery of sparse signals under noise. In the first part of our analysis, we have shown that in order for the OMP algorithm to accurately recover the support of any K-sparse signal, the SNR must be at least proportional to the sparsity level K of the signal. For high-dimensional setting, our result indicates that the exact support recovery with OMP is not possible under finite SNR.

In the second part of our analysis, we have considered a practical scenario where the SNR is an absolute constant independent of the sparsity K. While the exact support recovery with OMP is not possible for this scenario, our analysis has shown that recovery with an arbitrarily small but constant

fraction of errors is possible. For high-dimensional setting, our result offers an affirmative answer to the open question of whether OMP can perform the approximate support recovery of sparse signals with bounded SNR [12]. We would like to point out a technical limitation in this result. Unlike existing results for the exact support recovery that depend on the minimum magnitude of nonzero elements in the signal, our result for the approximate support recovery exhibits the dependence on the minimum as well as the maximum magnitudes (more precisely, the ratio κ). Deriving a similar result but without the dependence on the maximum magnitude would require a more refined analysis and our future work will be directed towards this avenue.

APPENDIX A
PROOF OF THEOREM 3.2

Proof: To prove the necessity of the lower bound of SNR in (12), it suffices to show that OMP may fail to recover the support of some sparse signal $\mathbf{x}$ when

$$\sqrt{\text{SNR}} \leq \frac{\sqrt{K}(1 + \delta_{K+1})}{(1 - \sqrt{K}\delta_{K+1}) \cdot \sqrt{\text{MAR}}}. \tag{46}$$

In the following, we will show that there exists a set of $\mathbf{\Phi}$, $\mathbf{x}$, and $\mathbf{v}$, for which (46) is satisfied but OMP fails to recover the support of $\mathbf{x}$.

Consider an identity matrix $\mathbf{\Phi}^{m \times m}$, a K-sparse signal $\mathbf{x} \in \mathcal{R}^m$ with all nonzero elements equal to one, and an 1-sparse noise vector $\mathbf{v} \in \mathcal{R}^m$ as follows,

$$\mathbf{\Phi} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ and } \mathbf{v} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Then the measurements are given by

$$\mathbf{y} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

In this case, we have $\delta_{K+1} = 0$,

$$\text{SNR} = \frac{\|\mathbf{\Phi x}\|_2^2}{\|\mathbf{v}\|_2^2} = K \text{ and } \text{MAR} = 1.$$

It is easily verified that condition (46) is satisfied; however, OMP fails to recover the support of $\mathbf{x}$. Specifically, OMP is not guaranteed to make a correct selection at the first iteration. ∎

APPENDIX B
PROOF OF PROPOSITION 3.3

Proof: We first give a proof of (24). Since u is the largest value in $\{|\langle \phi_i, \mathbf{r}^k \rangle|\}_{i \in \mathcal{T} \backslash \mathcal{T}^k}$,

$$
\begin{aligned}
u &= \max_{i \in \mathcal{T} \backslash \mathcal{T}^k} |\langle \phi_i, \mathbf{r}^k \rangle| \\
&\geq \frac{1}{\sqrt{|\mathcal{T} \backslash \mathcal{T}^k|}} \sqrt{\sum_{i \in \mathcal{T} \backslash \mathcal{T}^k} \langle \phi_i, \mathbf{r}^k \rangle^2} \\
&\overset{(a)}{=} \frac{1}{\sqrt{K - k}} \|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{r}^k\|_2 \\
&= \frac{1}{\sqrt{K - k}} \|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} (\mathbf{\Phi x} + \mathbf{v})\|_2 \\
&\overset{(b)}{\geq} \frac{1}{\sqrt{K - k}} \left(\|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi x}\|_2 - \|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{v}\|_2 \right),
\end{aligned}
\tag{47}
$$

where (a) is from (22) and (b) is due to the triangle inequality. Observe that

$$
\begin{aligned}
& \|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi x}\|_2 \\
& \overset{(a)}{=} \|\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2 \\
& \overset{(b)}{\geq} \frac{\|\mathbf{x}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2}{\|\mathbf{x}'_{\mathcal{T} \backslash \mathcal{T}^k}\|_2} \\
& \overset{(c)}{=} \frac{\|\mathbf{x}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} (\mathbf{P}^{\perp}_{\mathcal{T}^k})' \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2}{\|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2} \\
& = \frac{\|\mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2^2}{\|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2} \\
& \geq \frac{\lambda_{\min} \left((\mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k})' \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \right) \|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2^2}{\|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2} \\
& \overset{(d)}{=} \lambda_{\min} \left(\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \right) \|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2 \\
& \overset{(e)}{\geq} \lambda_{\min} (\mathbf{\Phi}'_{\mathcal{T} \cup \mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \cup \mathcal{T}^k}) \|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2 \\
& \overset{(f)}{\geq} (1 - \delta_K) \|\mathbf{x}_{\mathcal{T} \backslash \mathcal{T}^k}\|_2,
\end{aligned}
\tag{48}
$$

where (a) is because $\mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T}^k} = \mathbf{0}$, (b) is from the norm inequality, (c) and (d) use the fact that $\mathbf{P}^{\perp}_{\mathcal{T}^k} = (\mathbf{P}^{\perp}_{\mathcal{T}^k})' = (\mathbf{P}^{\perp}_{\mathcal{T}^k})^2$, (e) is from Lemma 2.5, and (f) is from the RIP. (Note that $\mathcal{T}^k \subset \mathcal{T}$ and so $|\mathcal{T} \cup \mathcal{T}^k| = |\mathcal{T}| = K$.) Also,

$$
\begin{aligned}
& \left\| \mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{v} \right\|_2 \\
& \overset{(a)}{=} \left\| (\mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k})' \mathbf{v} \right\|_2 \\
& \leq \sqrt{\lambda_{\max} \left((\mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k})' \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \right)} \|\mathbf{v}\|_2 \\
& \overset{(b)}{=} \sqrt{\lambda_{\max} \left(\mathbf{\Phi}'_{\mathcal{T} \backslash \mathcal{T}^k} \mathbf{P}^{\perp}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \backslash \mathcal{T}^k} \right)} \|\mathbf{v}\|_2 \\
& \overset{(c)}{\leq} \sqrt{\lambda_{\max} \left(\mathbf{\Phi}'_{\mathcal{T} \cup \mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T} \cup \mathcal{T}^k} \right)} \|\mathbf{v}\|_2 \\
& \leq \sqrt{1 + \delta_K} \|\mathbf{v}\|_2
\end{aligned}
\tag{49}
$$

where (a) and (b) are due to $\mathbf{P}^{\perp}_{\mathcal{T}^k} = (\mathbf{P}^{\perp}_{\mathcal{T}^k})' = (\mathbf{P}^{\perp}_{\mathcal{T}^k})^2$ and (c) is from Lemma 2.5.

Using (47), (48) and (49), we have

$$u \geq \frac{1}{\sqrt{K-k}} \left((1 - \delta_K) \|\mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k}\|_2 - \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right). \tag{50}$$

Next, we proceed to prove (24). Let $w \in \Omega$ denote the index corresponding to the largest element in $\{|\langle \phi_i, \mathbf{r}^k \rangle|\}_{i \in \Omega \setminus (\mathcal{T} \cup \mathcal{T}^k)}$. Then by the definition of v,

$$
\begin{aligned}
v &= |\langle \phi_w, \mathbf{r}^k \rangle| \\
&= |\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp (\mathbf{\Phi x} + \mathbf{v}) \rangle| \\
&\overset{(a)}{\leq} |\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi x} \rangle| + |\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{v} \rangle| \\
&\overset{(b)}{=} |\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi}_{\mathcal{T} \setminus \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k} \rangle| + |\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{v} \rangle|, \tag{51}
\end{aligned}
$$

where (a) is from the triangle inequality and (b) is because $\mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi}_{\mathcal{T}^k} = \mathbf{0}$.

Since $w \notin \mathcal{T}$ and also noting that $\mathcal{T}^k \subset \mathcal{T}$ (by the induction hypothesis), we have

$$|\{w\} \cup \mathcal{T}^k \cup (\mathcal{T} \setminus \mathcal{T}^k)| = |\{w\} \cup \mathcal{T}| = K + 1.$$

Using this together with Lemma 2.3, we have

$$
\begin{aligned}
&|\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi}_{\mathcal{T} \setminus \mathcal{T}^k} \mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k} \rangle| \\
&= \left\| \phi_w' (\mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi}_{\mathcal{T} \setminus \mathcal{T}^k}) \mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k} \right\|_2 \\
&\leq \delta_{|\{w\} \cup \mathcal{T}^k \cup (\mathcal{T} \setminus \mathcal{T}^k)|} \|\mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k}\|_2 \\
&= \delta_{K+1} \|\mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k}\|_2 \tag{52}
\end{aligned}
$$

and

$$
\begin{aligned}
|\langle \phi_w, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{v} \rangle| &= |\langle (\mathbf{P}_{\mathcal{T}^k}^\perp)' \phi_w, \mathbf{v} \rangle| \\
&\overset{(a)}{=} |\langle \mathbf{P}_{\mathcal{T}^k}^\perp \phi_w, \mathbf{v} \rangle| \\
&\overset{(b)}{\leq} \|\mathbf{P}_{\mathcal{T}^k}^\perp \phi_w\|_2 \|\mathbf{v}\|_2 \\
&\leq \|\phi_w\|_2 \|\mathbf{v}\|_2 \\
&\overset{(c)}{\leq} \sqrt{1 + \delta_1} \|\mathbf{v}\|_2, \tag{53}
\end{aligned}
$$

where (a) is due to the fact that $\mathbf{P}_{\mathcal{T}^k}^\perp = (\mathbf{P}_{\mathcal{T}^k}^\perp)'$, (b) is from Cauchy-Schwarz inequality, and (c) is from the RIP.

Finally, combining (51), (52), and (53) we have

$$v \leq \delta_{K+1} \|\mathbf{x}_{\mathcal{T} \setminus \mathcal{T}^k}\|_2 + \sqrt{1 + \delta_1} \|\mathbf{v}\|_2, \tag{54}$$

which completes the proof. ∎

APPENDIX C
PROOF OF (35)

Proof: First, from the definition of OMP (see Table I), we have that for any integer $k \geq 0$,

$$
\begin{aligned}
\mathbf{r}^k - \mathbf{r}^{k+1} &= (\mathbf{y} - \mathbf{\Phi x}^k) - (\mathbf{y} - \mathbf{\Phi x}^{k+1}) \\
&\overset{(a)}{=} \mathbf{P}_{\mathcal{T}^k} \mathbf{y} - \mathbf{P}_{\mathcal{T}^{k+1}} \mathbf{y} \\
&\overset{(b)}{=} (\mathbf{P}_{\mathcal{T}^{l+1}} - \mathbf{P}_{\mathcal{T}^{l+1}} \mathbf{P}_{\mathcal{T}^l}) \mathbf{y} \\
&= \mathbf{P}_{\mathcal{T}^{l+1}} \mathbf{P}_{\mathcal{T}^l}^\perp \mathbf{y} \\
&= \mathbf{P}_{\mathcal{T}^{l+1}} \mathbf{r}^k, \tag{55}
\end{aligned}
$$

where (a) is from that

$$\mathbf{x}^k = \underset{\mathbf{u}:supp(\mathbf{u})=\mathcal{T}^k}{\arg\min} \|\mathbf{y} - \mathbf{\Phi u}\|_2$$

and hence $\mathbf{\Phi x}^k = \mathbf{\Phi}_{\mathcal{T}^k} \mathbf{\Phi}_{\mathcal{T}^k}^\dagger \mathbf{y} = \mathbf{P}_{\mathcal{T}^l} \mathbf{y}$, (b) is because $span(\mathbf{\Phi}_{\mathcal{T}^k}) \subseteq span(\mathbf{\Phi}_{\mathcal{T}^{k+1}})$ so that $\mathbf{P}_{\mathcal{T}^k} \mathbf{y} = \mathbf{P}_{\mathcal{T}^{k+1}} (\mathbf{P}_{\mathcal{T}^k} \mathbf{y})$.

Since $span(\mathbf{\Phi}_{\mathcal{T}^{k+1}}) \supseteq span(\phi_{t^{k+1}})$, we have

$$\|\mathbf{r}^k - \mathbf{r}^{k+1}\|_2^2 = \|\mathbf{P}_{\mathcal{T}^{k+1}} \mathbf{r}^k\|_2^2 \geq \|\mathbf{P}_{t^{k+1}} \mathbf{r}^k\|_2^2.$$

Further, by noting that $\|\mathbf{r}^k - \mathbf{r}^{k+1}\|_2^2 = \|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2$, we have

$$
\begin{aligned}
\|\mathbf{r}^k\|_2^2 - \|\mathbf{r}^{k+1}\|_2^2 &\geq \|\mathbf{P}_{t^{k+1}} \mathbf{r}^k\|_2^2 \\
&\overset{(a)}{\geq} \|(\phi_{t^{k+1}}^\dagger)' \phi_{t^{k+1}}' \mathbf{r}^k\|_2^2 \\
&\overset{(b)}{=} (\phi_{t^{k+1}}' \mathbf{r}^k)^2 \|\phi_{t^{k+1}}^\dagger\|_2^2 \\
&\overset{(c)}{=} \|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2 \|\phi_{t^{k+1}}^\dagger\|_2^2 \\
&\overset{(d)}{\geq} \|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2 \left(\frac{\|\phi_{t^{k+1}}^\dagger \phi_{t^{k+1}}\|_2^2}{\|\phi_{t^{k+1}}\|_2^2} \right) \\
&\overset{(e)}{\geq} \frac{\|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2}{1 + \delta_1}, \tag{56}
\end{aligned}
$$

where (a) is because $\mathbf{P}_{t^{k+1}} = \mathbf{P}_{t^{k+1}}' = (\phi_{t^{k+1}}^\dagger)' \phi_{t^{k+1}}'$, (b) is from that $\phi_{t^{k+1}}' \mathbf{r}^k$ is a scalar, (c) is due to the identification rule of OMP, (d) is from the norm inequality, and (e) follows from the RIP. ∎

APPENDIX D
PROOF OF (36)

Proof: The main goal of the proof is to establish a lower bound on $\|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2$.

Observing that

$$(\mathbf{\Phi}' \mathbf{r}^k)_{\mathcal{T}^k} = \mathbf{\Phi}_{\mathcal{T}^k}' (\mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{y}) = (\mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{\Phi}_{\mathcal{T}^k})' \mathbf{y} = \mathbf{0},$$

we have

$$\|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2 = \|(\mathbf{\Phi}' \mathbf{r}^k)_{\Omega \setminus \mathcal{T}^k}\|_\infty^2 \tag{57}$$

Then it follows from the Hölder's inequality that for all $\mathbf{w} \in \mathcal{R}^n$,

$$
\begin{aligned}
\|\mathbf{\Phi}' \mathbf{r}^k\|_\infty^2 &\geq \frac{\langle (\mathbf{\Phi}' \mathbf{r}^k)_{\Omega \setminus \mathcal{T}^k}, \mathbf{w}_{\Omega \setminus \mathcal{T}^k} \rangle^2}{\|\mathbf{w}_{\Omega \setminus \mathcal{T}^k}\|_1^2} \\
&= \frac{\langle \mathbf{\Phi}' \mathbf{r}^k, \mathbf{w} \rangle^2}{\|\mathbf{w}_{\Omega \setminus \mathcal{T}^k}\|_1^2} \\
&= \frac{\langle \mathbf{r}^k, \mathbf{\Phi w} \rangle^2}{\|\mathbf{w}_{\Omega \setminus \mathcal{T}^k}\|_1^2} \tag{58}
\end{aligned}
$$

Let $\mathbf{w} \in \mathcal{R}^n$ be a vector such that

$$\mathbf{w}_{\mathcal{S}} = \begin{cases} \mathbf{x}_{\mathcal{S}} & \mathcal{S} \subseteq T \cap \mathcal{T}^k \cup \Gamma_\tau^k, \\ \mathbf{0} & \mathcal{S} \subseteq \Omega \setminus (T \cap \mathcal{T}^k \cup \Gamma_\tau^k). \end{cases} \tag{59}$$

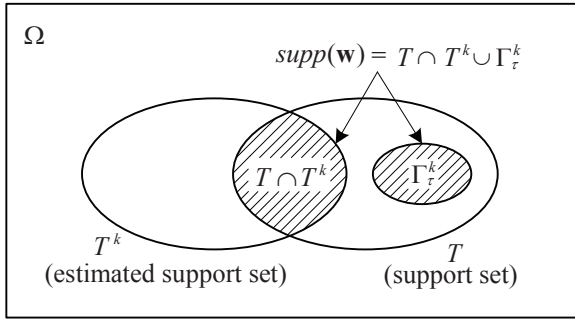where $\tau = \delta^{1/2}$. See Figure 1 for an illustration of $supp(\mathbf{w})$. Then (58) can be rewritten as

Fig. 1. Illustration of $supp(\mathbf{w})$.

$$
\begin{aligned}
\|\mathbf{\Phi}'\mathbf{r}^k\|_\infty^2 &\geq \frac{\langle \mathbf{\Phi}'\mathbf{r}^k, \mathbf{w}\rangle^2}{\|\mathbf{x}_{\Gamma_\tau^k}\|_1^2} \\
&\geq \frac{\langle \mathbf{r}^k, \mathbf{\Phi}\mathbf{w}\rangle^2}{\|\mathbf{x}_{\Gamma_\tau^k}\|_1^2} \\
&\overset{(a)}{\geq} \frac{\langle \mathbf{r}^k, \mathbf{\Phi}\mathbf{w}\rangle^2}{|\Gamma_\tau^k|\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2} \\
&= \frac{\langle \mathbf{r}^k, \mathbf{\Phi}\mathbf{w}\rangle^2}{\lceil \tau K\rceil\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2} \\
&\overset{(b)}{=} \frac{\langle \mathbf{r}^k, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2}{\lceil \tau K\rceil\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2}
\end{aligned} \tag{60}
$$

where (a) follows from the norm inequality ($\|\mathbf{u}\|_1 \leq \sqrt{\|\mathbf{u}\|_0}\,\|\mathbf{u}\|_2$) and (b) is because

$$
\langle \mathbf{r}^k, \mathbf{\Phi}\mathbf{x}^k\rangle = \langle \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{y}, \mathbf{P}_{\mathcal{T}^k}\mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{P}_{\mathcal{T}^k}^\perp \mathbf{P}_{\mathcal{T}^k}\mathbf{y}\rangle = \mathbf{0}.
$$

We now consider the term $\langle \mathbf{r}^k, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2$ in the numerator of the right-hand-side of (60). Since the residual can be expressed as $\mathbf{r}^k = \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k) + \mathbf{v}$, we have

$$
\begin{aligned}
&\langle \mathbf{r}^k, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 \\
&= \langle \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k) + \mathbf{v}, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 \\
&= (\langle \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k), \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle + \langle \mathbf{v}, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle)^2 \\
&\overset{(a)}{\geq} (1-\tau)\langle \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k), \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 - (1/\tau-1) \\
&\quad \times \langle \mathbf{v}, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 \\
&\overset{(b)}{\geq} (1-\tau)\langle \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k), \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 - (1/\tau-1) \\
&\quad \times \|\mathbf{v}\|_2^2\,\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\overset{(c)}{=} \frac{1-\tau}{4}\left(\|\mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k)\|_2^2 - \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x})\|_2^2 \right. \\
&\quad \left. + \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2\right)^2 - (1/\tau-1)\|\mathbf{v}\|_2^2\,\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2,
\end{aligned} \tag{61}
$$

where (a) uses the inequality

$$
(u+v)^2 \geq (1-\tau)u^2 - (1/\tau-1)\,v^2
$$

with $u = \langle \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k), \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle$ and $v = \langle \mathbf{v}, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle$, (b) is due to the Cauchy-Schwarz inequality, and (c) is from the fact that

$$
\langle \mathbf{u}, \mathbf{z}\rangle = \frac{1}{2}\left(\|\mathbf{u}\|_2^2 - \|\mathbf{u}-\mathbf{z}\|_2^2 + \|\mathbf{z}\|_2^2\right)
$$

with $\mathbf{u} = \mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k)$ and $\mathbf{z} = \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)$.

Further, observe that

$$
\begin{aligned}
&\|\mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k)\|_2^2 - \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x})\|_2^2 \\
&\overset{(a)}{=} \|\mathbf{\Phi}(\mathbf{x}-\mathbf{x}^k)\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k\backslash\Gamma_\tau^k}\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&\overset{(b)}{\geq} (1-\delta)\|\mathbf{x}-\mathbf{x}^k\|_2^2 - (1+\delta)\|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&\geq (1-\delta)\|(\mathbf{x}-\mathbf{x}^k)_{\Gamma^k}\|_2^2 - (1+\delta)\|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&\overset{(c)}{=} (1-\delta)\|\mathbf{x}_{\Gamma^k}\|_2^2 - (1+\delta)\|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&= (1-\delta)\left(\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2 + \|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2\right) \\
&\quad - (1+\delta)\|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&= (1-\delta)\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2 - 2\delta\|\mathbf{x}_{\Gamma^k\backslash\Gamma_\tau^k}\|_2^2 \\
&\overset{(d)}{\geq} (1-\delta)\lceil\tau K\rceil(x_\tau^k)^2 - 2\delta K(x_\tau^k)^2 \\
&\geq \left(1-\delta - \frac{2\delta}{\tau}\right)\lceil\tau K\rceil(x_\tau^k)^2 \\
&\overset{(e)}{\geq} (1-3\tau)\lceil\tau K\rceil(x_\tau^k)^2,
\end{aligned} \tag{62}
$$

where (a) is from (59), (b) is due to the RIP, (c) is because $\mathbf{x}_{\Gamma^k}^k = \mathbf{0}$, (d) is because x_τ^k is the $\lceil\tau K\rceil$-th largest element (in magnitude) in $\mathbf{x}_{\Gamma^k}$ and hence is the smallest one in $\mathbf{x}_{\Gamma_\tau^k}$, and (e) is from that

$$
1 - \delta - \frac{2\delta}{\tau} = 1 - \delta - 2\tau \geq 1 - 3\tau.
$$

Using (61) and (62), we have

$$
\begin{aligned}
&\langle \mathbf{r}^k, \mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\rangle^2 \\
&\geq \frac{1-\tau}{4}\left((1-3\tau)\lceil\tau K\rceil(x_\tau^k)^2 + \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2\right)^2 \\
&\quad - (1/\tau-1)\|\mathbf{v}\|_2^2\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\overset{(a)}{\geq} (1-\tau)\left((1-3\tau)\lceil\tau K\rceil(x_\tau^k)^2\right)\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\quad - (1/\tau-1)\|\mathbf{v}\|_2^2\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\overset{(b)}{\geq} \left((1-4\tau)\lceil\tau K\rceil(x_\tau^k)^2 - (1/\tau-1)\|\mathbf{v}\|_2^2\right) \\
&\quad \times \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\overset{(c)}{\geq} (1-5\tau)\lceil\tau K\rceil(x_\tau^k)^2\|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2 \\
&\overset{(d)}{\geq} (1-5\tau)\lceil\tau K\rceil(x_\tau^k)^2(1-\delta)\|\mathbf{w}-\mathbf{x}^k\|_2^2 \\
&\overset{(e)}{\geq} (1-6\tau)\lceil\tau K\rceil(x_\tau^k)^2\|\mathbf{w}-\mathbf{x}^k\|_2^2 \\
&\geq (1-6\tau)\lceil\tau K\rceil(x_\tau^k)^2\|(\mathbf{w}-\mathbf{x}^k)_{\Omega\backslash\mathcal{T}^k}\|_2^2 \\
&= (1-6\tau)\lceil\tau K\rceil(x_\tau^k)^2\|\mathbf{x}_{\Gamma_\tau^k}\|_2^2
\end{aligned} \tag{63}
$$

where (a) follows from the fact that $(u+v)^2 \geq 4uv$ with $u = (1-3\tau)\lceil\tau K\rceil(x_\tau^k)^2$ and $v = \|\mathbf{\Phi}(\mathbf{w}-\mathbf{x}^k)\|_2^2$, (b) is because

$$
(1-\tau)(1-3\tau) = 1 - 4\tau + 3\tau^2 \geq 1 - 4\tau,
$$

(c) is because 1) the assumption $\text{SNR} \geq \kappa^2 \delta^{-3/2}$ implies that

$$
\begin{aligned}
\|\mathbf{\Phi x}\|_2^2 &= \text{SNR} \cdot \|\mathbf{v}\|_2^2 \\
&\geq \frac{\kappa^2}{\delta^{3/2}} \|\mathbf{v}\|_2^2 \\
&= \frac{\kappa^2}{\tau^3} \|\mathbf{v}\|_2^2 \\
&\geq \frac{\kappa^2}{\tau^3} \left(1 - \tau \left(\tau - \frac{1}{2} \right)^2 - \frac{3\tau}{4} \right) \|\mathbf{v}\|_2^2 \\
&= \frac{\kappa^2(1+\tau^2)(1-\tau)\|\mathbf{v}\|_2^2}{\tau^3},
\end{aligned}
\tag{64}
$$

and 2) on the other hand,

$$
\begin{aligned}
\|\mathbf{\Phi x}\|_2^2 &\overset{\text{RIP}}{\leq} (1+\delta)\|\mathbf{x}\|_2^2 \\
&\leq (1+\delta)K x_{\max}^2 \\
&\leq (1+\delta)K(\kappa x_{\min})^2 \\
&\leq (1+\delta)K(\kappa x_\tau^k)^2 \\
&= (1+\tau^2)K(\kappa x_\tau^k)^2,
\end{aligned}
\tag{65}
$$

and hence

$$
(1/\tau - 1)\|\mathbf{v}\|_2^2 \leq \tau^2 K (x_\tau^k)^2 \leq \tau \lceil \tau K \rceil (x_\tau^k)^2,
\tag{66}
$$

(d) is from the RIP, and (e) follows from that

$$
(1-5\tau)(1-\delta) \geq (1-5\tau)(1-\tau) = 1 - 6\tau + 5\tau^2 \geq 1 - 6\tau.
$$

Finally, using (60) and (63), we have

$$
\begin{aligned}
\|\mathbf{\Phi}'\mathbf{r}^k\|_\infty^2 &\geq \frac{\langle \mathbf{r}^k, \mathbf{\Phi}(\mathbf{w} - \mathbf{x}^k)\rangle^2}{\lceil \tau K \rceil \|\mathbf{x}_{\Gamma_\tau^k}\|_2^2} \\
&\geq \frac{(1-6\tau)\lceil \tau K \rceil (x_\tau^k)^2 \|\mathbf{x}_{\Gamma_\tau^k}\|_2^2}{\lceil \tau K \rceil \|\mathbf{x}_{\Gamma_\tau^k}\|_2^2} \\
&= (1-6\tau)(x_\tau^k)^2,
\end{aligned}
\tag{67}
$$

which completes the proof. ∎

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] R. A. DeVore and G. G. Lorentz, *Constructive approximation*, vol. 303, Springer, 1993.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1999.

[5] R. L. Schaefer, "Subset selection in regression," *Technometrics*, vol. 34, no. 2, pp. 229–229, 1992.

[6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*. IEEE, Nov. Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[9] T. Zhang, "On the consistency of feature selection using greedy least squares regression," *J. of Mach. Learn. Res.*, vol. 10, pp. 555–568, 2009.

[10] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[11] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[12] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit: A brownian motion analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1010–1021, Mar. 2012.

[13] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[14] M. A. Davenport and M. B. Wakin, "Analysis of Orthogonal Matching Pursuit using the restricted isometry property," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[15] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.

[16] S. Huang and J. Zhu, "Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP," *Inverse Problems*, vol. 27, no. 3, pp. 035003, 2011.

[17] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.

[18] E. D. Livshits, "On the efficiency of the orthogonal matching pursuit in compressed sensing," *Sbornik: Mathematics*, vol. 203, no. 2, pp. 183, 2012.

[19] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, Sep. 2012.

[20] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit algorithm," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3654–3656, Jun. 2012.

[21] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.

[22] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[23] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[24] J. Wen, X. Zhu, and D. Li, "Improved bounds on restricted isometry constant for orthogonal matching pursuit," *Electronics Letters*, vol. 49, no. 23, pp. 1487–1489, 2013.

[25] L. Chang and J. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 60, no. 9, pp. 5702–5715, Sep. 2014.

[26] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[27] Y. Shen and S. Li, "Sparse signals recovery from noisy measurements by orthogonal matching pursuit," *arXiv:1105.6177*, 2011.

[28] R. Wu, W. Huang, and D Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 403–406, Apr. 2013.

[29] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, Mar. 2009.

[30] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

[31] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[32] Y. Shen, B. Li, W. Pan, and J. Li, "Analysis of generalised orthogonal matching pursuit using restricted isometry constant," *Electronics Letters*, vol. 50, no. 14, pp. 1020–1022, 2014.

[33] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[34] S. Foucart, "Stability and robustness of weak orthogonal matching pursuits," in *Recent Advances in Harmonic Analysis and Applications*, pp. 395–405. Springer, 2013.

[35] J. Wang and B. Shim, "How many iterations are needed for the exact recovery of sparse signals using orthogonal matching pursuit," *arXiv:1211.4293v2*, 2012.

[36] E. D. Livshitz and V. N. Temlyakov, "Sparse approximation and recovery by greedy algorithms," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 3989–4000, Jul. 2014.

[37] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[38] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[39] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[40] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[41] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.

[42] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[43] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.